

WHAT IS CLAIMED IS:

1. A method for automatically determining at least one modal value of non-numeric data comprises:
selecting a data subset from a dataset, the data subset comprising at least a
5 portion of the dataset and including at least one non-numeric value; and
automatically determining at least one modal value based on the selected data subset.
2. The method of Claim 1, wherein selecting the data subset from the
10 dataset comprises querying a database.
3. The method of Claim 1, each value of the data subset comprising one of the following data types:
float;
15 integer;
currency;
date;
decimal; or
string.
20
4. The method of Claim 1, wherein determining at least one modal value based on the selected data subset comprises:
sorting the selected data subset by value;
processing the sorted data subset to identify one or more modal groups, each
25 modal group comprising one or more instances of a substantially identical value; and
determining at least one modal value based, at least in part, on the one or more modal groups.

5. The method of Claim 4 further comprising determining a modal count for each modal group, each modal count comprising the number of instances of the substantially identical value in the associated modal group.

5 6. The method of Claim 5, wherein determining at least one modal value based, at least in part, on the one or more modal groups comprises:
 determining a highest one or more modal counts;
 selecting the substantially identical value from each modal group associated with the highest model count; and
10 assigning each selected substantially identical value to one modal value.

7. The method of Claim 5, in response at least in part to each modal count being equal to one, assigning a null value to one modal value.

15 8. The method of Claim 4, one of the modal groups comprising at least one lowercase string value and at least one mixed-case string value.

9. The method of Claim 1, wherein determining at least one modal value based on the selected data subset comprises:
20 selecting one data object from the data subset;
 comparing a value of the data object to a plurality of stored values in a lookup table, each stored value being associated with one modal count;
 in response, at least in part, to the value of the data object being located in the plurality of stored values, adding one to the associated modal count;
25 selecting the highest one or more modal counts from the lookup table; and
 assigning each stored value associated with one of the highest modal counts to one modal value.

10. Software for automatically determining at least one modal value of non-numeric data operable to:

select a data subset from a dataset, the data subset comprising at least a portion of the data set and including at least one non-numeric value; and

5 automatically determine at least one modal value based on the selected data subset.

11. The software of Claim 10, wherein the software operable to select the data subset of the dataset comprises software operable to query a database.

10

12. The software of Claim 10, each value of the data subset comprising one of the following data types:

float;

integer;

15 currency;

date;

decimal; or

string.

20 13. The software of Claim 10, wherein the software operable to determine at least one modal value based on the selected data subset comprises software operable to:

sort the selected data subset by value;

25 process the sorted data subset to identify one or more modal groups, each modal group comprising one or more instances of a substantially identical value; and

automatically determine at least one modal value based, at least in part, on the one or more modal groups.

14. The software of Claim 13 further operable to determine a modal count for each modal group, each modal count comprising the number of instances of the substantially identical value in the associated modal group.

5 15. The software of Claim 14, wherein the software operable to determine at least one modal value based, at least in part, on the one or more modal groups comprises software operable to:
 determine a highest one or more modal counts;
 select the substantially identical value from each modal group associated with
10 the highest model count; and
 assign each selected substantially identical value to one modal value.

16. The software of Claim 14, in response at least in part to each modal count being equal to one, further operable to assign a null value to one modal value.

15

17. The software of Claim 13, one of the modal groups comprising at least one lowercase string value and at least one mixed-case string value.

18. The software of Claim 10, wherein the software operable to determine at least one modal value based on the selected data subset comprises software operable to:

selecting one data object from the data subset;

5 comparing a value of the data object to a plurality of stored values in a lookup table, each stored value being associated with one modal count;

in response to the value of the data object being located in the plurality of stored values, adding one to the associated modal count;

selecting the highest one or more modal counts from the lookup table; and

10 assigning each stored value associated with one of the highest modal counts to one modal value.

19. System for automatically determining at least one modal value of non-numeric data comprises:

memory operable to store a data set, the data set comprising a plurality of data objects and each data object comprising a data type and a value; and

5 one or more processors operable to:

select a data subset from the dataset, the data subset comprising at least a portion of the plurality of data objects and including at least one non-numeric data object; and

10 automatically determine at least one modal value based on the selected data subset.

20. The system of Claim 19, wherein the processors operable to select the data subset of the dataset comprise processors operable to query a database.

15 21. The system of Claim 19, each data object comprising one of the following data types:

float;
integer;
currency;
20 date;
decimal; or
string.

22. The system of Claim 19, wherein the processors operable to determine at least one modal value based on the selected data subset comprise processors operable to:

sort the selected data subset by value;

5 process the sorted data subset to identify one or more modal groups, each modal group comprising two or more instances of a substantially identical value; and automatically determine at least one modal value based, at least in part, on the one or more modal groups.

10 23. The system of Claim 22, the processors further operable to determine a modal count for each modal group, each modal count comprising the number of instances of the substantially identical value in the associated modal group.

15 24. The system of Claim 23, wherein the processors operable to determine at least one modal value based, at least in part, on the one or more modal groups comprise processors operable to:

determine a highest one or more modal counts;

select the substantially identical value from each modal group associated with the highest model count; and

20 assign each selected substantially identical value to one modal value.

25 25. The system of Claim 23, in response at least in part to each modal count being equal to one, the processors further operable to assign a null value to one modal value.

26. The system of Claim 22, one of the modal groups comprising at least one lowercase string value and at least one mixed-case string value.

27. The system of Claim 19, wherein the processors operable to determine at least one modal value based on the selected data subset comprise processors operable to:

selecting one data object from the data subset;

5 comparing a value of the data object to a plurality of stored values in a lookup table, each stored value being associated with one modal count;

in response to the value of the data object being located in the plurality of stored values, adding one to the associated modal count;

selecting the highest one or more modal counts from the lookup table; and

10 assigning each stored value associated with one of the highest modal counts to one modal value.

28. A system for automatically determining at least one modal value of non-numeric data comprises:

means for selecting a data subset from a dataset, the data subset comprising at least a portion of the dataset and including at least one non-numeric value; and

5 means for automatically determining at least one modal value based on the selected data subset.